EL-GY 6143: Introduction to Machine Learning
(Fall 2024, Updated 8/22/2024)

Overview
This course provides a hands-on approach to machine learning and statistical pattern recognition. The course will describe fundamental algorithms for linear regression, classification, model selection, support vector machines, neural networks, dimensionality reduction and clustering. The course includes demos and labs on real and synthetic data using Python. Applications will be demonstrated in audio and image processing, robotic control, gene expression data, neural decoding, and text processing. No prior machine learning experience is required.

Students will learn to: Formulate problems using a variety of simple ML models. Use software packages to train and validate models. Analyze the performance of these methods using tools from optimization and probability. Pre-process data and visualize results from various sources (time series, audio, image, text, etc.).

Course instructor: Dr. Pei Liu (peiliu at nyu.edu)

Lead TA: Meghana Kuchana (kk4347 at nyu.edu)

Graders: TBD (at nyu.edu)

Lectures: Wednesday 5:00 PM - 7:30 PM, Jacobs Academic Bldg, Room 474

Office Hour:

All office hours are online via Zoom. Check Brightspace for time and Zoom link. If you want to meet in-person, please email us.

Course websites:

Most up-to-date copy of lecture notes, demos and labs and HWs organized in course units are available at: https://github.com/pliugithub/MachineLearning/blob/master/sequence.md

Textbooks:

- Hastie, Tibshirani, Friedman, "Elements of Statistical Learning". http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf
- Raschka, "Python Machine Learning", 2015. http://file.allitebooks.com/20151017/Python%20Machine%20Learning.pdf https://sebastianraschka.com/books.html

Supplementary texts and resources
- Bishop, "Pattern Recognition and Machine Learning"
- James, Witten, Hastie and Tibshirani, "An Introduction to Statistical Learning", http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf
- Installing python (need to do this before first recitation): http://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html
- Python tutorial: https://docs.python.org/3/tutorial/

Grading:

- Midterm 30%, Final 30%, Labs and homework 20%, Project 20% (up to two people).
- Project is OPTIONAL. If you choose not to do it, I will use the higher grade of your midterm and final as your project grade.

- Midterm and final exams are closed book with cheat sheets. Students will need to be able to write simple python code on paper in the exams.
- HWs, labs and their due date are posted on GitHub. No late submission accepted except under extraordinary circumstances and must be approved in advance by the instructor. HW and Labs are to be hand-in on Gradescope.
- When submitting labs, please
  - Restart python kernel and cleaning output
  - Re-run all cells and make sure it runs
  - Submit both .ipynb and .pdf files

Pre-requisites:

- Undergraduate level probability and linear algebra. No ML experience is expected for this class.
- Students may NOT enroll in this class if they have taken any one of: CSE-GY 6923 (Grad Intro ML), EE-UY 4563 (UG Intro ML), EL-GY 9133 (Grad Advanced ML).
- Students with prior ML experience are encouraged to take graduate-level Probability (EL-GY 6303) in the Fall and advanced ML in the Spring.
- Programming experience is essential, including some exposure or willingness to learn object-oriented programming. No experience in python is required as python will be taught as part of the class. If you are familiar with Java, C/C++, you should fine with learning Python.

Tentative Course Schedule

These materials are developed by Prof. Sundeep Rangan, and I will be updating during the semester.

Week 1 (9/4): Introduction to machine learning: Examples, types of ML problems. Course logistics. Intro to python and jupyter and GitHub. Single variable linear regression (Unit 1) (Brief introduction).

HW: Students should download python, jupyter, github, and do the lab and HW in Unit 1.

Week 2 (9/11): Simple linear regression (Unit2): Linear models, least squares formula; Extensions for non-linear models.

Week 3 (9/18): Multiple Linear Regression (Unit 3).

Week 4 (9/25): Model selection and regularization (Unit 4): Understanding underfitting and overfitting with polynomial fitting; Irreducible error due to measurement noise; Bias and variance tradeoff; Cross validation.

Week 5 (10/2): Lasso and regularization (Unit 5).

Week 6 (10/9): Logistic regression and classification (Unit 6).

Week 7 (10/16): Midterm

Week 8 (10/23): Numerical optimization (Unit 7): Unconstrained optimization, gradient descent, global vs. local minima, convexity. Example with logistic regression. Implementation with Python (Possibly move some of the notation of tensor and gradient with respect to tensor in Unit 7 here).

Week 9 (10/30): Support vector machines (Unit 8): Image classification; SVM formulation, support vectors; Duality, kernel methods.

Week 10 (11/6): Neural networks (Unit 9): Formulation, motivation; Computation graphs, backpropagation; Introduction to tensorflow and keras; Stochastic gradient descent.

Week 11 (11/13): Convolutional and deep networks (Unit 10): Convolutional layers, Pooling, batch normalization; Advanced Tensorflow features; Using GPUs.

Week 12 (11/20): Dimensionality reduction (Unit 11): Principal component analysis (unsupervised), linear discriminant analysis (supervised), LDA SVDs

Week 13 (11/27): Unsupervised Clustering (Unit 12): K-means, Mixture models, EM methods

Week 14 (12/4): Tree based methods (Unit 13): Decision tree, Random Forest, Boosting.

Week 15 (12/11): No class(Legislative Friday).

Week 15 (12/18): Final exam, (Optional project is due on the end of the day of 12/22).

If you are experiencing an illness or any other situation that might affect your academic performance in a class, please email Deanna Rayment, Coordinator of Student Advocacy, Compliance and Student Affairs.

Deanna can reach out to your instructors on your behalf when warranted.

- deanna.rayment@nyu.edu
- https://engineering.nyu.edu/staff/deanna-rayment

Inclusion Statement:

The NYU Tandon School values an inclusive and equitable environment for all our students. I hope to foster a sense of community in this class and consider it a place where individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations, and abilities will be treated with respect. It is my intent that all students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. If this standard is not being upheld, please feel free to speak with me.